



AI Governance

An overview to planning and implementing AI Governance driven transformation.

September 18, 2026





George Marountas

**Director, AI Assurance –
Internal Audit & Controls**

George Marountas is a Technology Director at KPMG with 22 years of experience leading multimillion-dollar global IT transformations. George specializes in developing continuous risk management programs that leverage advanced analytics and tools to detect and mitigate information security risks. He currently directs a diverse client portfolio, leading AI Governance Audits, Sarbanes-Oxley (SOX) compliance programs, and Cyber Threat Intelligence initiatives to ensure robust internal controls and drive demonstrable results.

Agenda

01 Why AI Governance Now

The risks, the urgency, and what the board is asking

02 Building the Governance Program

Program layers and the enterprise and use-level governance layers

03 AI Governance Control Considerations

System cards, AI-specific controls, and an illustrative audit

04 Practical Steps for Internal Audit

Where to start and what to do next



01 | Why AI Governance Now

Why this matters now

The value of using AI has driven implementation and use. AI is arriving faster than governance models were designed to handle.



"AI can be confidently wrong, easily manipulated, or deployed outside formal oversight channels." – Source: COSO, "Achieving Effective Internal Control Over Generative AI," February 2026

91% of leaders say data security, privacy and risk will influence AI strategy over the next six months.

35% of leaders cited “Lack of risk and governance processes” as one of the greatest risks / barriers to successfully meeting GenAI strategy goals.

Boards are increasingly covering AI topics on a quarterly basis, including:
48% - Risk management processes / governance
45% - Trust in the accuracy and fairness of outputs.

Source: KPMG Q3 2025- Q1 2026 AI Pulse Surveys



© 2026 KPMG LLP, a Delaware limited liability partnership, and its subsidiaries are part of the KPMG global organization of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved. USCS038586-1A



The Board Is Asking Questions That Most Execs Can't Yet Fully Answer

"How many AI models are in production — and who approved them?"

Most organizations can't produce an AI inventory

"What happens if our AI makes a biased decision?"

Liability exposure without documented controls

"Are we compliant with the EU AI Act?"

Regulatory timelines are accelerating

"Who is accountable for AI risk?"

Ownership is often unclear or fragmented

Setting the Context – Why AI Governance is Different

Internal governance models are experiencing a paradigm shift, as traditional human & machine governance models now require re-invention to account for the emergence of agentic AI in an organization's operating model.



Traditional Governance

Human / Process Governance

- Internal controls over humans rely on **supervision, review, and training.**
- Many human-driven processes are not scalable, but their work can be **sampled and validated through established governance processes.**

Machine Governance

- Machines perform **recurring tasks the same way every time** (e.g. processing a million invoices identically)
- Users are able to audit machines through **deterministic checks**, expecting the same output each time a process is run

AI-Driven Governance

AI Considerations

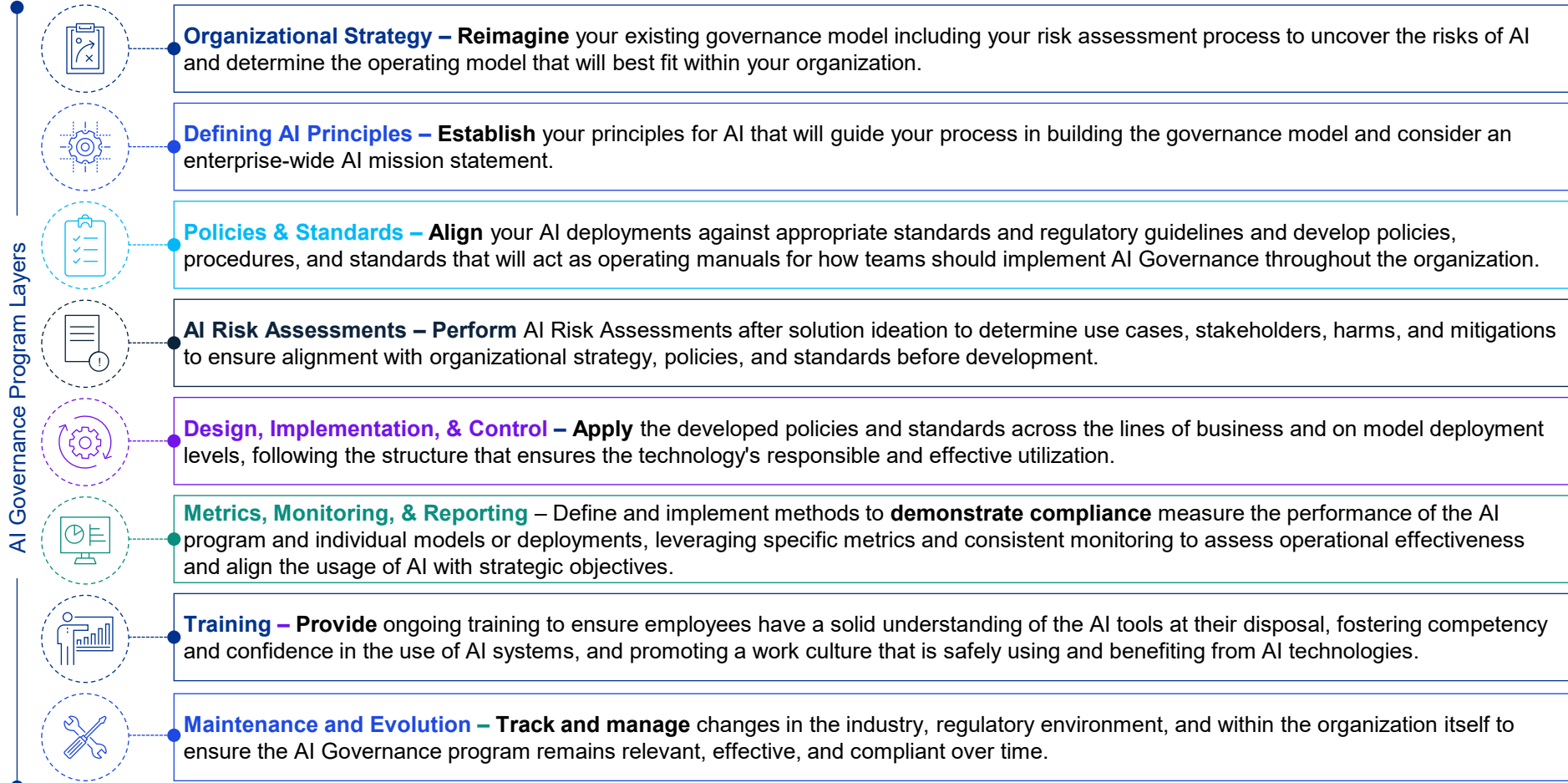
Problem Statement: How will an organization **govern, supervise, and validate GEN or Agentic AI** when it combines the scale of machines and the potential unpredictability of humans?

- Agents can operate at a machine's scale, but like humans, are not perfect and can **introduce probabilistic outcomes** that may deviate from the expected / intended results
- Organizations are looking for **assurance in the reliability of their AI**
- There is a need to **re-invent traditional governance models to account for audits of its AI agents' outputs**



02 | Building the Governance Program

AI Governance Program Layers



These program layers will...

-  Be a baseline to compare against your current AI governance program
-  Provide a contextualized viewpoint of the future state across the layers
-  Balance the needs of the function with the abilities of the technology and organization
-  Demystify the technology and highlight how it can support the functional objectives
-  Explore the impact of business change through every layer of the organization

While enterprise governance might be mature, rapid adoption now raises a use-level question that the enterprise layer was never built to answer alone.

MATURE ENTERPRISE GOVERNANCE IN PLACE

- ✓ Enterprise risk management and risk appetite discipline
- ✓ Model risk management under supervisory guidance
- ✓ Cybersecurity and technology risk governance
- ✓ Vendor / third-party risk management
- ✓ An established assessment process for new technology and AI

WHAT THAT LAYER ANSWERS

Is the platform, architecture, and infrastructure sound?

Are models validated and monitored?

Is data protected; are vendors controlled?

Are enterprise operational risks within appetite?

THE QUESTION VOLUME NOW RAISES

Hundreds of heterogeneous uses on the same few approved platforms

Risk set by context-of-use: who consumes the output, in which process, with what oversight

A use-level question the enterprise layer was never designed to answer alone

This is about scaling for growth rather than fixing what is broken.

Oversight should scale with autonomy and risk. Each step up keeps every control below it and adds its own.

TIER 1

Traditional AI / ML

Predictive models, scoring

MRM lifecycle): validation,
monitoring, inventory

TIER 2

Generative AI

Content, summarization, copilots

+ use-tiering, output validation,
prompt/guardrail controls, usage
monitoring

TIER 3

Agentic AI

Multi-step tasks, tool use

+ human in/on-the-loop design,
tool invocation limits, kill
switches, action logging

TIER 4

Orchestrated agents

Multi-agent workflows

+ agent-to-agent controls, non-
human identity governance,
containment & isolation

INCREASING AUTONOMY →

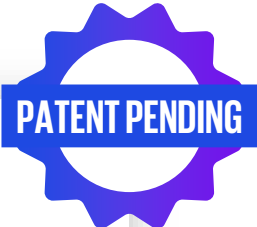
Controls **add up**: each tier keeps everything to its left and adds its own.



03 | AI Governance Control Considerations

System Card approach for testing and evaluating AI systems and usages

AI system cards provide a single place to find key information about AI systems such as intended use, limitations, data considerations, and performance against responsible frameworks.

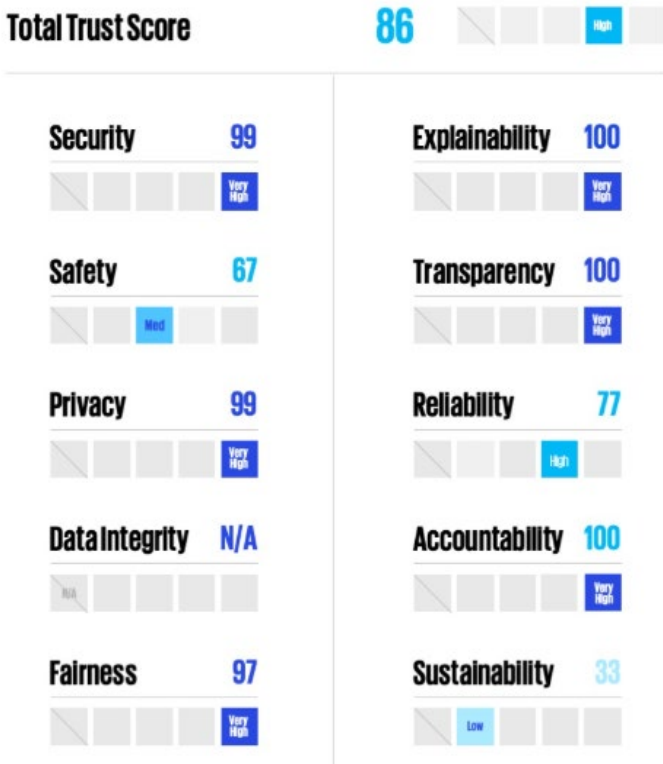


What is an “AI System”



An AI system infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy & adaptiveness. (OECD- 2023)

Sample output – Trusted AI score



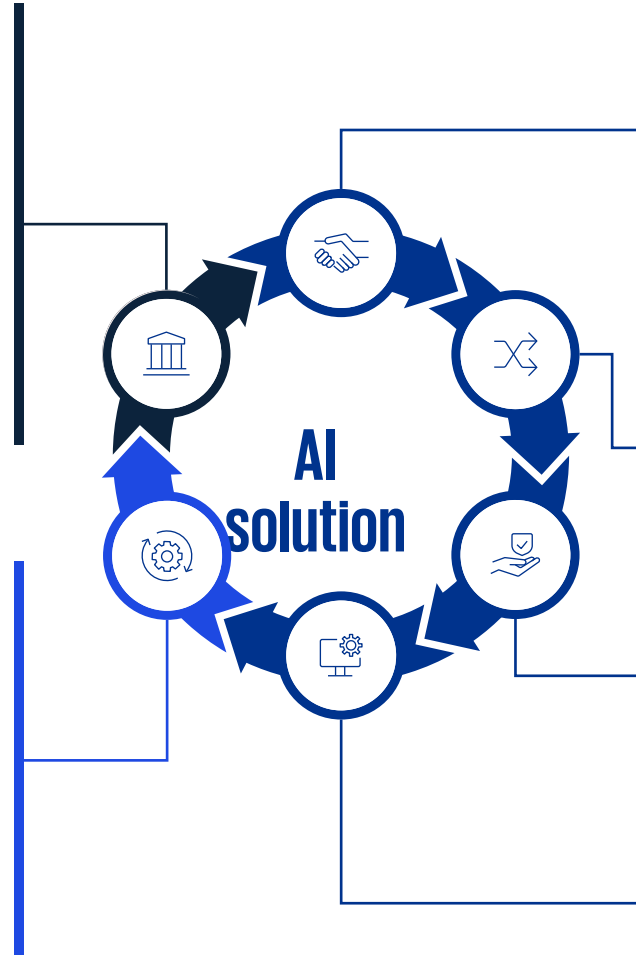
AI-specific control considerations

Governance/Entity level controls

- Governance policies and procedures around the use of AI
- Aligning AI's use to the entity's financial reporting objectives
- Policies and procedures over monitoring models, testing, and compliance
- Managing and monitoring relationships with third parties including service organizations for AI use
- Involvement and oversight by the Board

Business processes/controls

- **Explainability:** entity's ability to understand how and why a conclusion was made by AI
- **Reliability:** ability of AI to consistently operate with its intended purpose upon implementation and over time
- **Data quality:** accuracy, completeness, and appropriateness of the data used by the AI model



General IT Controls

Program acquisition and development

- Documentation of algorithm design, decision-making processes, and assumptions
- Various test cases and conditions to evaluate algorithm performance and identify potential issues
- Data Validation processes for training and test data used in AI model

Program Change

- Establishing process to update and retrain AI model as needed to adapt to new data or changing requirements

Access to Programs and Data

- Limit access of AI model to only programs and data within their intended scope
- Restrict 3rd party access to entity's sensitive data
- Establish clear guidelines for human involvement/access to AI output

Computer operations

- Set up monitoring mechanisms to notify relevant stakeholders when model drift is detected

Example – Illustrative Audit GenAI User Access Review



Data Extraction and Ingestion

Data Extraction: GenAI extracts user counts and entitlement totals from source systems (e.g., Active Directory, ERPs).

Audit Objective: ensure user lists and entitlement data from all in-scope applications are completely and accurately extracted from the source.



Data Transformation & Integration

Data Standardization: GenAI correlates user identities across different applications (e.g., j.smith vs. john.smith).

Audit Objective: verify user identities and entitlements from different systems are correctly normalized and correlated into a single view.



Judgment & Forecasting

Ruleset: GenAI's configured ruleset correctly identifies assigned roles, permissions and flags SoD conflicts.

Audit Objective: assess the accuracy and completeness of the AI's risk analysis (e.g., SoD violations, dormant accounts, excessive privileges).



Workflow Orchestration

Review tasks: documented workflow for routing review tasks (e.g., routing a user's access profile to their direct manager) function correctly.

Audit Objective: confirm the access review and certification workflow operates as designed.



Automated Transaction Processing

Reconciliation Report: GenAI reconciles the final certification report with the detailed sign-offs recorded from the reviews.

Audit Objective: ensure the results of the review (certifications, required changes) are accurately recorded and actioned.



Monitoring & Continuous Review

Analysis: GenAI performs continuous SoD analysis and periodic reviews as scheduled.

Audit Objective: validate that the system continuously monitors for new conflicts and that the review process is completed in a timely manner.



Knowledge Retrieval

User inquiries: Inquire GenAI's "explainability" feature by asking it to justify why a certain mix of access rights can be considered a conflict.

Audit Objective: ensure that data presented to reviewers is complete and concise to enable informed decisions.



Human-AI Collaboration

Clarity: Managers understand easily what they are approving when they interact with the GenAI's user interface to certify access.

Audit Objective: facilitate an effective review to capture an audit trail of human decisions.



04 | Practical Steps for Internal Audit

AI: Practical steps for Internal Audit

Below, we provide illustrative steps that combine the perspectives of 1) AI as an auditable area and 2) AI as a tool for the IA function:



Key Session Takeaways



Technology, automation and AI Governance will continue to have a profound impact on the design and execution of processes and controls



Internal Audit will play a key role in governing AI and redefining control activities and requirements to fully leverage the possibilities



Questions & Answers



George Marountas
Director, AI Assurance

Email: gmarountas@kpmg.com | Cell: 703-966-3665

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

Learn about us:



kpmg.com

The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act upon such information without appropriate professional advice after a thorough examination of the particular situation.

© 2026 KPMG LLP, a Delaware limited liability partnership, and its subsidiaries are part of the KPMG global organization of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved. USCS038586-1A

The KPMG name and logo are trademarks used under license by the independent member firms of the KPMG global organization.



Appendix 1 – Supporting Frameworks and Case Examples

Leading AI Governance Frameworks

A quick comparison of the leading AI frameworks: **NIST AI RMF** is a flexible, lifecycle-based method for technical risk management; **ISO 42001** is a formal, certifiable management system for organizational governance; and **COSO** builds AI oversight into existing internal control and enterprise risk structures.

	NIST AI RMF RISK METHODOLOGY	ISO 42001 MANAGEMENT SYSTEM	COSO AI GUIDANCE
Type	Voluntary guidance, self-directed, developed by a U.S. government agency.	Certifiable standard audited by accredited bodies; voluntary and international.	Guidance that extends existing enterprise risk and internal control structures to include AI.
Primary purpose	Risk identification and mitigation.	Enterprise-wide governance and risk management.	Anchors AI within enterprise risk and control so outputs are reliable and systems are “audit-ready.”
Scope	Org-defined (single system or broad); applies across all industries and AI lifecycle stages.	Organization-wide, full AI lifecycle.	Internal control over AI-enabled processes and outputs.
Certifiability	Voluntary, non-certifiable framework.	Certifiable standard.	Not certifiable; principles-based guidance.
Core structure	Govern – Map – Measure – Manage	Plan – Do – Check – Act (management model)	Five components mapped to AI risks: Control Environment, Risk Assessment, Control Activities, Info & Communication, Monitoring.
Regulatory alignment	Encourages risk-aware practices; non-prescriptive.	Supports alignment with formal regulations (e.g., EU AI Act).	Supports internal control and financial reporting expectations (e.g., ICFR).



COSO: Achieving Effective Internal Control Over GenAI

RELEASED FEB 23, 2026

COSO 2013 Internal Control

Integrated Framework, extended for Generative AI.

Background

On February 23, 2026, COSO released a roadmap with a practical, COSO-aligned path to use GenAI responsibly.

Features

Adds GenAI-specific practices for each of the **17 principles** in the COSO 2013 Internal Control — Integrated Framework.

Tailors risk and control considerations to **eight distinct GenAI capability types**.

Provides examples for all capability types, including **minimum control expectations** for all five COSO components.

17

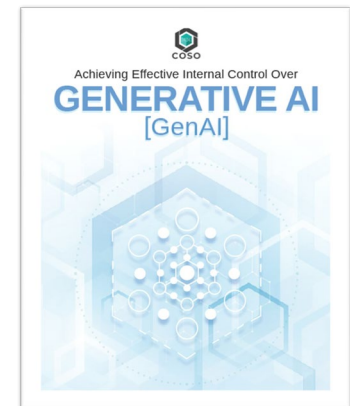
Principles

8

Capability types

5

Components



AI Operations & Observability Metrics for AI Systems

Example observability metrics implemented within to support GenAI Ops monitoring.

Performance & Quality	Resource Efficiency	Risk & Safety	Observability & Reliability
Measures how well the AI model or agent meets intended outcomes in terms of accuracy, speed, and uptime.	Measure how efficiently AI systems use infrastructure resources such as compute, memory, performance and cost.	Measure how often AI systems produce unsafe, biased, or untrustworthy outputs and how well they resist attacks and follow policies.	Monitor how well AI systems are traced and observed to ensure reliability, catch issues early, and maintain trust.
User Metrics Average user star rating % of chatbot sessions completed without user giving up	GPU Utilization % of GPU compute in use	Hallucination Rate % of outputs containing unsupported or false information	Continuous Evaluation Scores Accuracy or quality from live test queries in production
Evaluation Error Rate % of outputs that fail quality checks or evaluation tests	CPU & Memory Usage % of CPU utilized and RAM footprint of the AI service	Bias/Unfairness Incidents Count of outputs flagged for biased or derogatory content	Error/Exception Rate Frequency of runtime errors across requests (crashes, invalid outputs)
Precision & Recall % of content produced is correct/ on target	Token Usage per Request Number of tokens processed for each query (input & output)	Toxic Content Rate Frequency of harmful content generated by category	Trace Coverage % of requests fully traced and logged.
Latency & Throughput Average response time Requests per second (RPS) queries or API calls per second	Cost per Token/Query \$ cost/token or query	Adversarial Attack Success Prompt injection success rate	Model Drift (Decay) Trend in model performance on a fixed reference test set
Reliability & Uptime % of time service is available Number of failed requests/exceptions	Total AI Spend Aggregate cost of running GenAI system	Compliance Violations Count of times AI violated internal or regulatory policies	Anomalies Detected Count of anomalies flagged by monitoring in a period
	Disk I/O & Storage Read/write throughput and storage usage		Alert Response Times Average minutes to acknowledge an alert



Agentic AI Evaluation Benchmarks

Example observability metrics specific to Agentic AI helps to ensure sufficient coverage and monitoring specific to the unique emerging risks agentic brings.

Planning Reliability	Tool Precision	Human Collaboration	Trusted
Measures how well the agent breaks down and follows multi-step tasks.	Measures accuracy and effectiveness of tool selection	Measures effectiveness of oversight mechanisms	Measures other key attributes aligned to Trusted AI pillars
Knowledge Reliability Degree to which the provided context enables accurate and relevant answers	Tool Selection Accuracy % of times the agent chooses right tool for task	Human-in-the-loop compliance Degree to which the agent adheres to human-in-the-loop policies during operation	Policy Alignment % of responses compliant with defined policies
Intent Understanding Degree to which the agent correctly interprets the user's intent	# of API calls Number of calls to external APIs during task completion	Human intervention # of requests for human intervention per goal	Refusal Rate % of unsafe or out-of-scope queries correctly declined
Agent Success Rate % of successfully completed tasks	# of Interactions between agents Number of requests for human intervention per tasks	Self-aware failure rate # of successful failures where agent is aware of its limitations in solving a task and requests support	Safety # of safety incidents/overall interactions
Instruction Adherence Degree to which intermediate steps followed instructions		Oversight Compliance % of required checkpoints correctly handled	Latency per task/goal Overall time to per task and per goal
Goal Adherence Degree to which output followed provided instructions		Transparency Quality Quality of explanations provided for human review	Other Evaluations: Bias, Hallucination, Data Privacy, Security

Process control activities - Example control considerations



Explainability:

The entity's ability to understand how and why a conclusion was made by AI.

Controls may vary depending on whether the entity has access to the underlying algorithms

- The inputs that the model relies on to execute a task or decision are documented and reviewed by a manager.
- The documentation maintained on how the model functions and makes its decisions is reviewed and re-evaluated quarterly.
- Reports from model interpretability techniques (methods to gain insights into how the model makes decisions) or user feedback analysis are reviewed to evaluate the ongoing explainability of the model.



Reliability of the model:

The ability of AI to consistently operate with its intended purpose upon implementation and overtime.

Controls may vary depending on whether the AI employs static models or models that could potentially change over time and impact the entity's use

- Errors or discrepancies generated by the AI are logged and reviewed by a manager for further investigation and resolution and to determine if updates to the model are necessary.
- The decisions underlying the design of the model are reviewed to determine if there are changes to the entity or its business that impact the design or functionality.
- For third-party models that are developed outside the entity, the manager evaluates the technology is operating effectively for its intended purpose. The accounting manager reviews statements released by third parties that explain any updates to the model or algorithms, if applicable.

Case Scenario: Healthcare Organization

ILLUSTRATIVE — REPRESENTATIVE ENGAGEMENT PATTERN

Situation

Large health system deploying AI for clinical decision support, claims processing, and patient scheduling. No centralized AI inventory. Board concerned after media coverage of AI bias in healthcare.

What We Found

27+ AI systems in production (12 previously unknown to IT), no formal model validation process, HIPAA implications unaddressed for AI-processed PHI, vendor AI terms insufficient.

What We Built

AI risk classification tailored to healthcare (clinical vs. operational vs. administrative), governance committee charter, AI-specific audit program mapped to HIPAA and emerging FDA guidance.

Outcome

CAE presented first AI assurance report to audit committee within 6 months. Proactive regulatory posture. Three high-risk AI systems identified for enhanced monitoring before any adverse events.

Case Scenario: Financial Services Firm

ILLUSTRATIVE — REPRESENTATIVE ENGAGEMENT PATTERN

Situation

Mid-market bank using AI for credit decisioning, fraud detection, and customer service chatbot. OCC exam approaching. Existing model risk management framework doesn't address AI/ML models adequately.

What We Found

Model validation focused on traditional statistical models — ML models lacked explainability documentation. No bias testing for protected classes in lending. SOX implications for AI-generated financial data unaddressed.

What We Built

Enhanced MRM framework incorporating AI-specific validation requirements, fair lending AI testing protocol, SOX AI control matrix, regulatory-ready documentation templates.

Outcome

OCC examination completed with no AI-related findings. Proactive fair lending compliance. Board gained visibility into AI risk portfolio for the first time. Audit function positioned as strategic advisor.

Case Scenario: Manufacturing Enterprise

ILLUSTRATIVE — REPRESENTATIVE ENGAGEMENT PATTERN

Situation

Global manufacturer with AI in predictive maintenance, quality control, and supply chain optimization. EU operations subject to AI Act. No visibility into vendor-embedded AI across production systems.

What We Found

40+ AI touchpoints across ERP, MES, and IoT platforms — most vendor-embedded and undocumented. Safety-critical AI in quality inspection without formal validation. EU AI Act high-risk classification triggered for worker safety applications.

What We Built

Cross-functional AI governance committee, vendor AI assessment methodology, safety-critical AI validation protocol, EU AI Act compliance roadmap with phased implementation.

Outcome

Complete visibility into AI landscape for the first time. Safety-critical systems prioritized for enhanced controls. EU compliance roadmap accepted by legal and presented to supervisory board.



Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

Learn about us:



kpmg.com

The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act upon such information without appropriate professional advice after a thorough examination of the particular situation.

© 2026 KPMG LLP, a Delaware limited liability partnership, and its subsidiaries are part of the KPMG global organization of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved.

The KPMG name and logo are trademarks used under license by the independent member firms of the KPMG global organization.